From the Departments of
OPERATIONS AND MANAGEMENT SCIENCE / MARKETING

>

Bayesian Estimation of Circumplex Models Subject to Prior Theory Constraints and Scale-Usage Bias

by

Peter Lenk

ASSOCIATE PROFESSOR OF OPERATIONS
AND MANAGEMENT SCIENCE

Michel Wedel

DWIGHT F. BENTON PROFESSOR
OF MARKETING

Ulf Böckenholt

PROFESSEUR BELL OF MARKETING/
BELL PROFESSOR OF E-MARKETING AT
MCGILL UNIVERSITY

This paper presents a hierarchical Bayes circumplex model for ordinal ratings data. The circumplex model was proposed to represent the circular ordering of items in psychological testing by imposing inequalities on the correlations of the items. We provide a specification of the circumplex, propose identifying constraints and conjugate priors for the angular parameters, and accommodate theory-driven constraints in the form of inequalities. We investigate the performance of the proposed MCMC algorithm and apply the model to the analysis of value priorities data obtained from a representative sample of Dutch citizens.

Bayesian Estimation of Circumplex Models Subject to Prior Theory Constraints and Scale-Usage Bias[1]

Peter Lenk, [2] **Michel Wedel** and **Ulf Böckenholt**

The University of Michigan and McGill University

Abstract

This paper presents a hierarchical Bayes circumplex model for ordinal ratings data. The circumplex model was proposed to represent the circular ordering of items in psychological testing by imposing inequalities on the correlations of the items. We provide a specification of the circumplex, propose identifying constraints and conjugate priors for the angular parameters, and accommodate theory-driven constraints in the form of inequalities.We investigate the performance of the proposed MCMC algorithm and apply the model to the analysis of value priorities data obtained from a representative sample of Dutch citizens.

1 Introduction

A classical finding in psychometrics is that similarity judgements of different colors can be represented in a two-dimensional space in the form of Newton's color circle (Shepard, 1962a, 1962b). Based on this work, similar circular representations proved useful for describing variations among experiences or judgments in a wide range of psychological and related disciplines. For example, affective states are commonly depicted by a circular structure based on the dimensions of valence and arousal (Russell and Carroll, 1999). Numerous other applications can be found in personality and social psychology (Lippe 1995 and Plutchik & Conte

---

[1] We wish to thank Michael Browne and two anonymous reviewers for their comments.

[2] The University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109-1234. Phone 734-936-2619 & email plenk@umich.edu

1997). The circular ordering of the responses implies that the elements of the corresponding correlation matrix follow a so–called circumplex structure with correlations first decreasing but then increasing as one moves diagonally from the main diagonal. Guttman (1954) and Anderson (1960) suggested stochastic processes on the perimeter of the circle that produce positive correlations obeying the circumplex structure, with, respectively, moving average and Markov properties. Models that allow for negative correlations were developed by Cudeck (1986) and Wiggins, Steiger and Gaelick (1981). Browne (1992) proposed an extension of Anderson's (1960) model that allows for negative correlations.

We extend the work by Browne (1992) as follows. First, we introduce a Bayesian specification of the circumplex for ratings data and present identifying constraints and conjugate priors for the angular parameters. Second, we specify inequality constraints on blocks of variables in the circumplex as defined by psychological theories. Third, we accommodate idiosyncratic response-scale usage by persons (see for example, Rossi, Gilula and Allenby 2001) that, if not accounted for, may substantially distort the derived circumplex. Since surveys are often burdened with item non–response, we also capitalize on the MCMC estimation algorithm to impute missing values. As a result, our approach facilitates powerful tests of psychological theories based on a circumplex structure and controls for a number of nuisance effects.

The next section describes the proposed model and discusses estimation and inference issues. In Section 3, the model is fitted to value ratings collected in the Netherlands, and it is investigated whether the underlying data structure is consistent with the prominent circumplex value theory of Schwartz and Bilsky (1987, 1990). Section 4 summarizes the paper.

2   Model

Subject i responds $W_{i,j}$ to item j on an rating scale with H ordered categories. Such scales are very common in psychology and other social science applications. We assume that the

observed response is driven by a latent variable $Y_{i,j}$ falling between two cutpoints:

$$w_{i,j} = k \text{ iff } c_{i,k-1} < y_{i,j} < c_{i,k} \text{ for } k = 1, \ldots, H$$

where the cutpoints $\{c_{i,k}\}$ may vary from person to person. We consider it important to estimate person–specific cutpoints, since response scale bias has been reported to be highly idiosyncratic (Rossi, Gilula and Allenby 2001). The cutpoints are ordered: $c_{i,k-1} < c_{i,k}$, and the first two and last two cutpoints are fixed without loss of generality: $c_{i,0} = -\infty$, $c_{i,1} = -1$, $c_{i,H-1} = 1$, $c_{i,H} = \infty$. The probability of the ordinal response is:

$$Pr(W_{i,j} = k) = \int_{c_{i,k-1}}^{c_{i,k}} f(y_{i,j}) dy \text{ for } k = 1, \ldots, H \tag{1}$$

where f is the density of $Y_{i,j}$.

Browne (1992) proposed using trigonometric series to model the circumplex correlations between items and developed a corresponding factor analytical model. We similarly specify a random effects model for the latent response variable $Y_{i,j}$ to describe individual differences for person i and item j:

$$Y_{i,j} = \mu_j + \phi_i + \alpha_i \sin(\theta_j) + \beta_i \cos(\theta_j) + \epsilon_{i,j} \tag{2}$$
$$\text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, J; \; \theta_1 = 0; \text{ and } 0 \leq \theta_j < 2\pi$$

The mean, latent response for item j is μ_j; ϕ_i is a subject-specific random effect that captures scale–usage effects, and $\alpha_i \sin(\theta_j) + \beta_i \cos(\theta_j)$ is a subject by item interaction term that provides circumplex correlations. The error terms, $\{\epsilon_{i,j}\}$, are mutually independent, normally distributed random error terms with zero mean, and item-specific variances: $\text{var}(\epsilon_{i,j}) = \sigma_j^2$.

The model can be viewed as a three–factor model where ϕ_i, α_i, and β_i are subject–specific factor scores. The first factor score ϕ_i is a random effect that takes into account subject–specific scale usage effects. These effects are artifacts of the measurement system and usually do not have substantive, field-dependent implications: they reflect that subjects use systematically different parts of the ordinal measurement scale. Respondents with a positive ϕ_i tend to use the upper end of the rating scale, and respondents with a negative

ϕ_i tend to use the lower end. We will see in the application that ignoring scale–usage effects can severely distort the estimated circumplex.

The other two factor scores α_i and β_i are individual–level random coefficients that have substantive meaning for the psychological phenomenon under investigation: they represent bipolar latent constructs. Their item-specific loadings, $\sin(\theta_j)$ and $\cos(\theta_j)$, are constrained to the unit circle; thus, they are expressed in polar coordinates. With the appropriate assumptions about these random effects and the constrained loading, inter-item correlations, after adjusting for scale-usage bias, have a circumplex structure.

We assume that the random effects, $(\phi_i, \alpha_i, \beta_i)$, are mutually independent and normally distributed with zero means and the following variances: $\text{var}(\phi_i) = \lambda^2$; $\text{var}(\alpha_i) = \tau_a^2$; and $\text{var}(\beta_i) = \tau_b^2$. Circumplex correlations are obtained when $\tau_a = \tau_b = \tau$. If these variances are unequal, then one can reparameterize the subject by item interactions as $\alpha_i \tau_a \sin(\theta_j) + \beta_i \tau_b \cos(\theta_j)$ were α_i and β_j are factor scores with mean 0 and variance 1. Then the loadings $\tau_a \sin(\theta_j)$ and $\tau_b \cos(\theta_j)$ are constrained to the ellipse. In the empirical application, we will compare the circumplex model to this more general one.

After integrating out the random-effects, the variance and covariances of the latent variables for circumplex correlations $(\tau_a = \tau_b = \tau)$ for the items conditional on the angles are:

$$
\begin{aligned}
\text{var}(Y_{i,j}) &= \lambda^2 + \tau^2[\sin(\theta_j)^2 + \cos(\theta_j)^2] + \sigma_j^2 &(3) \\
&= \lambda^2 + \tau^2 + \sigma_j^2 \\
\text{cov}(Y_{i,j}, Y_{i,k}) &= \lambda^2 + \tau^2[\sin(\theta_j)\sin(\theta_k) + \cos(\theta_j)\cos(\theta_k)] &(4) \\
&= \lambda^2 + \tau^2 \cos(\theta_j - \theta_k)
\end{aligned}
$$

This covariance is a special case of Browne's (1992) approximation using first order, trigonometric polynomials. In the classical analysis of random-effects models, the variances and covariances in Equations (3) and (4) determine the error covariance matrix in the log-likelihood function. In Bayesian inference, the random effects $(\phi_i, \alpha_i, \beta_i)$ are frequently treated as unknown parameters that are estimable: they are not just nuisance parameters.

2.1 Identifiability

The part of the circumplex correlation function that depends on item angles, $\tau^2 \cos(\theta_j - \theta_k)$ from Equation (4), depends only on the differences in the angles so that the origin is arbitrary. Thus, we fix θ_1 to 0, but this alone does not identify the model, which can be seen as follows. Define another set of angles as $\psi_1 = 0$ and $\psi_j = 2\pi - \theta_j$ for $j \geq 2$. Because $\sin(2\pi - \theta)$ $= -\sin(\theta)$ and $\cos(2\pi - \theta) = \cos(\theta)$, the likelihoods $L[\alpha, \beta, \theta]$ and $L[-\alpha, \beta, \psi]$ are equal. Consequently, we identify the model by introducing a second constraint: $0 < \theta_2 < \pi$. Thus, θ_1 locates the origin, and θ_2 determines the positive direction.

We selected the cosine function in the covariance terms to represent the circumplex, but other functions (see Browne 1992) could be used as well. A function satisfies the circumplex properties for correlations if it is even, continuous, monotonically decreasing on $(0, \pi)$, monotonically increasing on $(\pi, 2\pi)$ with maxima of 1 at 0 and 2π, and a minimum of -1 at π. We choose the cosine without loss of generality, however, since the angles and trigonometric function are simultaneously unidentified for a finite set of items. If another function f has the circumplex properties, then it is possible to define a new set of angles ϕ such that the (ϕ, f) and $(\theta, \cos)$ result in the same covariances and likelihoods for a finite set of items. However, strictly speaking the invariance only holds with respect to the likelihood. For the posterior distributions of the parameters for the two models defined through (ϕ, f) and $(\theta, \cos)$ to be equivalent, the prior for θ would have to transformed into an equivalent prior for ϕ. In practice, since the prior specification is often chosen as a compromise between realism and convenience, the transformed prior for ϕ would rarely match a preferred direct specification of it. For example, an uninformative prior for θ, may not result in an uninformative prior for ϕ. [3]

---

[3] We thank one anonymous reviewer for pointing this out to us.

2.2 Block Constraints on the Angles

Substantive theory often postulates that subsets of angles in the circumplex, characterizing the items in a certain domain, are less than or greater than other subsets of angles in other domains, thus imposing blocks of constraints on the directions $\{\theta_j\}$. In psychological theories on personality and value priorities, such domains, consisting of groups of substantively homogeneous items, are often distinguished. Suppose that there are K blocks of constraints, and let $\mathcal{B}_k$ be the set of indices for the kth block. We are interested in specifying prior constraints on the order of these blocks of angles. Without loss of generality we assume the blocks are ordered from $1, ..., K$ with $\theta_1 = 0$ and θ_2 belonging to the first block.

Expressing the block constraints is fairly straightforward, except for the fact that angles in the first block can be on both sides of the origin. For blocks 2 to K, define the minimum and maximum angles:

$$\underline{\mathcal{B}_k} = \min\{\theta_j : j \in \mathcal{B}_k\} \text{ and } \overline{\mathcal{B}_k} = \max\{\theta_j : j \in \mathcal{B}_k\} \text{ for } k = 2, \ldots, K.$$

The "minimum" and "maximum" angles for the first block require some care because angles in the first block can be on both sides of the zero value:

$$
\begin{aligned}
\underline{\mathcal{B}_1} &= \begin{cases} 2\pi & \text{if } \underline{\mathcal{B}_K} > \max\{\theta_j : j \in \mathcal{B}_1\} \\ \min\{\theta_j : j \in \mathcal{B}_1 \text{ and } \theta_j > \overline{\mathcal{B}_K}\} & \text{if } \underline{\mathcal{B}_K} < \max\{\theta_j : j \in \mathcal{B}_1\} \end{cases} \\
\overline{\mathcal{B}_1} &= \max\{\theta_j : j \in \mathcal{B}_1 \text{ and } \theta_j < \underline{\mathcal{B}_2}\}
\end{aligned}
$$

That is, $\underline{\mathcal{B}_1}$ is the smallest angle in the first block that is larger than the angles in the Kth block. If no angle satisfies this requirement, it is defined to be 2π. $\overline{\mathcal{B}_1}$ is the largest angle in the first block that is smaller than the angles in the second block. With this nonstandard definition of "minimum" and "maximum" for the first block, we obtain the ordering:

$$\overline{\mathcal{B}_1} < \underline{\mathcal{B}_2} < \overline{\mathcal{B}_2} < \cdots < \underline{\mathcal{B}_K} < \overline{\mathcal{B}_K} < \underline{\mathcal{B}_1}.$$

These constraints are additional to $\theta_1 = 0$ and $0 < \theta_2 < \pi$.

2.3 Distributions

Introducing matrix notation simplifies defining the circumplex model in Equation (2) and its distributions. The $n \times J$ matrix for the latent scores $\{y_{i,j}\}$ (subject i and item j) from the cutpoint model is:

$$Y = \begin{bmatrix} y_{1,1} & \cdots & y_{1,J} \\ \vdots & \ddots & \vdots \\ y_{n,1} & \cdots & y_{n,J} \end{bmatrix} = \begin{bmatrix} \underline{Y}'_{1,\bullet} \\ \vdots \\ \underline{Y}'_{n,\bullet} \end{bmatrix} = [\underline{Y}_{\bullet,1} \ldots \underline{Y}_{\bullet,J}] \text{ where}$$

$$\underline{Y}_{i,\bullet} = J \text{ vector of latent item scores for subject } i;$$

$$\underline{Y}_{\bullet,j} = n \text{ vector of latent subject scores for item } j.$$

The $n \times J$ matrix of error terms is:

$$E = \begin{bmatrix} \epsilon_{1,1} & \cdots & \epsilon_{1,J} \\ \vdots & \ddots & \vdots \\ \epsilon_{n,1} & \cdots & \epsilon_{n,J} \end{bmatrix} = \begin{bmatrix} \underline{\epsilon}'_1 \\ \vdots \\ \underline{\epsilon}'_n \end{bmatrix}.$$

The item specific J–vectors for means and angles are:

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_J \end{bmatrix} ; \underline{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_J \end{bmatrix} ; \underline{X}_S = \begin{bmatrix} \sin(\theta_1) \\ \vdots \\ \sin(\theta_J) \end{bmatrix} \text{ and } \underline{X}_C = \begin{bmatrix} \cos(\theta_1) \\ \vdots \\ \cos(\theta_J) \end{bmatrix}.$$

The $J \times 2$ factor loading matrix is represented as:

$$X = [\underline{X}_S \, \underline{X}_C].$$

The $J \times J$ diagonal matrix of error variances is:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_J^2 \end{bmatrix}.$$

Factor scores and scale-usage effects are collected into n–vectors:

$$\underline{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} ; \underline{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} ; \text{ and } \underline{\phi} = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_n \end{bmatrix}.$$

With these definitions, Equation (2) becomes:

$$Y = \underline{\phi}\underline{1}'_J + \underline{1}_n\underline{\mu}' + \underline{\alpha}X'_S + \underline{\beta}\underline{X}'_C + E$$

where $\underline{1}_K$ is a K vector of ones.

We use the bracket notation "[•]" of Gelfand and Smith (1990) to designate a distribution or density for a random variable. The argument in the brackets identifies the distribution; for example, $[X]$ and $[Y|X]$ are the distributions of X and Y given X, respectively. The model and analysis require four distributions: uniform, normal, inverted gamma, and univariate extended Von Mises. The densities for the first three, standard distributions are displayed below to establish notation:

$$[v|a,b] = U(v|a,b) = (b-a)^{-1} \text{ for } a \leq v \leq b,$$

$$[\underline{x}|\underline{\mu},\Sigma] = N_m(\underline{x}|\underline{\mu},\Sigma) = (2\pi)^{-\frac{1}{2}}|\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})\right],$$

$$[y|a,b] = IG(y|a,b) = \frac{b^a}{\Gamma(a)}y^{-(a+1)}\exp(-b/y) \text{ for } y > 0.$$

The density for the extended Von Mises distribution is:

$$[\theta|\underline{d},Q,\mathcal{C}] = VM(\theta|\underline{d},Q,\mathcal{C}) \propto \exp\left\{-\frac{1}{2}[\underline{\xi}(\theta)-\underline{d}]'Q[\underline{\xi}(\theta)-\underline{d}]\right\}\chi(\theta \in \mathcal{C}),$$

where $\underline{\xi}(\theta) = [\sin(\theta),\cos(\theta)]'$; $\underline{d}$ is a 2 dimensional vector; Q is a 2 by 2 matrix; $\chi(\bullet)$ is the indicator function, and $\mathcal{C}$ is a subset of $[0,2\pi)$. Q does not need to be symmetric or positive definite because the range of ξ is finite. Figure 1 graphs the extended Von Mises distribution for $\underline{d} = (\sin(\pi),\cos(\pi))'$ and $Q = 5I_2$ when $\mathcal{C} = [0,2\pi]$. If Q is a matrix of zeros the density is uniform. If $\underline{d} = (\sin(\tilde{\theta}),\cos(\tilde{\theta}))'$, then the mode of the distribution is $\tilde{\theta}$. We will restrict our attention to quadratic forms because the likelihood function for the angles takes this expression, and the extended Von Mises is the natural conjugate prior distribution for the angles.

[INSERT FIGURE 1 ABOUT HERE]

8

The error terms of the latent variables are mutually independent and normally distributed:

$$[\underline{\epsilon}_i | \Sigma] = N_J(\underline{\epsilon}_i | \underline{0}_J, \Sigma) \text{ where } \underline{0}_J \text{ is a } J \text{ vector of zeros.}$$

The random effects are also mutually independent and normally distributed:

$$[\underline{\phi} | \lambda^2] = N_n(\underline{\phi} | \underline{0}_n, \lambda^2 I_n); \quad [\underline{\alpha} | \tau^2] = N_n(\underline{\alpha} | \underline{0}_n, \tau^2 I_n); \text{ and } [\underline{\beta} | \tau^2] = N_n(\underline{\beta} | \underline{0}_n, \tau^2 I_n)$$

where I_n is the $n \times n$ identity matrix. Given the latent variables and cutpoints, the distribution of the response for subject i is:

$$Pr(w_{i,1}, \ldots, w_{i,J} | c_i, \mu, \phi_i, \alpha_i, \beta_i, \theta, \Sigma)$$

$$= \prod_{j=1}^{J} \left[\int_{c_{i,w_{i,j}-1}}^{c_{i,w_{i,j}}} N_1[y_{i,j} | \mu_j + \phi_i + \alpha_i \sin(\theta_j) + \beta_i \cos(\theta_j), \sigma_j^2] dy_{i,j} \right].$$

The prior distribution for the cutpoints is conditionally uniform:

$$[c_i] \propto \chi(-1 < c_{i,2} < \ldots < c_{i,H-2} < 1).$$

That is, given $c_{i,k-1}$ and $c_{i,k+1}$, the conditional distribution of $c_{i,k}$ is uniform:

$$[c_{i,k} | c_{i,k-1}, c_{i,k+1}] = U(c_{i,k} | c_{i,k-1}, c_{i,k+1}) \text{ for } k = 2, \ldots, H-2.$$

The mean latent scores have a normal prior:

$$[\underline{\mu}] = N_J(\underline{\mu} | \underline{m}_0, V_0),$$

and the error and random effects variances have inverted gamma distributions:

$$[\sigma_j^2] = IG\left(\sigma_j^2 \Big| \frac{r_0}{2}, \frac{s_0}{2}\right); \quad [\lambda^2] = IG\left(\lambda^2 \Big| \frac{u_{0,1}}{2}, \frac{v_{0,1}}{2}\right); \text{ and } [\tau^2] = IG\left(\tau^2 \Big| \frac{u_{0,2}}{2}, \frac{v_{0,2}}{2}\right).$$

The prior distribution for the angles is extended Von Mises:

$$[\theta_j] = \begin{cases} VM(\theta_j | \underline{d}_0, Q_0, [0, \pi]) & \text{for } j = 2 \\ VM(\theta_j | \underline{d}_0, Q_0, [0, 2\pi)) & \text{for } j > 2. \end{cases}$$

with $\underline{d}_0 = (0, 1)'$, $Q_0 = 0.2 I_2$ where I_2 is the 2×2 identity matrix. This prior distribution is fairly flat on $[0, 2\pi)$.

2.4   MCMC Estimation

A primary goal of Bayesian inference is to compute the posterior distribution of the unknown parameters given the data. The posterior distribution quantifies the uncertainty about unknown parameters after observing the data. The posterior mean is the Bayes estimator under squared–error loss, and the posterior standard deviation is a measure of uncertainty about the parameter. For sufficiently large samples and well-behaved models, posterior distributions are approximately normal, and there is approximately 95% probability that the true parameter is within $\pm$ two posterior standard deviations of the posterior mean. For non-Bayesians, the posterior mean is the point estimator, and the posterior standard deviation roughly resembles the standard error for the point estimator. Estimation of the model is accomplished via Markov chain Monte Carlo (MCMC) (c.f. Gelfand and Smith 1990). The appendix gives details for the application to the circumplex model. After an initial transition period, the random deviates from MCMC can be treated as random draws from the posterior distribution and used to numerically approximate posterior statistics of the parameters. For example, the posterior mean is approximated by the average of the random draws.

The accuracy of these numerical approximations can be ascertained by the root mean squared simulation error (RMSSE). The RMSSE is the standard deviation of the MCMC approximation to the posterior mean and accounts for the autocorrelation in the Markov chain. The RMSSE tends to decrease as one uses more iterations in MCMC. In comparison, the theoretical posterior standard deviation does not depend on the estimation algorithm, and it tends to decrease as sample sizes increase. It is important to differentiate between the posterior standard deviation and the RMSSE. The first quantifies the posterior uncertainty about a parameter, while the latter quantifies the accuracy of the numerical algorithm in approximating the posterior mean. We will report the RMSSE to give an indication of the accuracy of the numerical approximations from the MCMC algorithm.

2.5 Brier Scores

Brier (1950) proposed a squared error loss statistic that compares predictive probabilities and random outcomes (c.f. Gordon and Lenk 1991, 1992). Let $\{z_i\}$ be n uncertain events where $z_i = 1$ if the event occurs and $z_i = 0$ otherwise. The Brier score is $BS = n^{-1} \sum_{i=1}^{n} (z_i - p_i)^2$ where p_i is the predictive probability for z_i. Suppose that one uses m different predictive probabilities $\{q_j\}$. The Brier score can be decomposed into two components, called "calibration" and "refinement":

$$BS = \frac{1}{n} \sum_{j=1}^{m} \sum_{i=1}^{n} \chi(p_i = q_j)(z_i - q_j)^2$$

$$= \frac{1}{n} \left[\sum_{j=1}^{m} n_j(q_j - \hat{q}_j)^2 + \sum_{j=1}^{m} \sum_{i=1}^{n} \chi(p_i = q_j)(z_i - \hat{q}_j)^2 \right]$$

$$n_j = \sum_{i=1}^{n} \chi(p_i = q_j) \text{ and } \hat{q}_j = n_j^{-1} \sum_{i=1}^{n} \chi(p_i = q_j) z_i$$

where n_j is the number of times that q_j is used; and $\hat{q}_j$ is the relative frequency of event j given that one predicted it would happen with probability q_j.

The first term of the decomposition is calibration and is related to bias. The calibration measure is zero when the predictive probability and conditional relative frequencies are equal. Clearly, calibration alone does not imply an accurate forecasting system. For instance, a system is well–calibrated if it always reports the base rates for events; however, base rates may not be very informative. The second measure, refinement, is similar to variance, and measures the propensity of the prediction system to use values close to zero or one: in a well-calibrated system, forecasts closer to zero or one are more useful than forecasts in the middle of the unit interval. DeGroot and Feinberg (1982) showed that if two systems are well-calibrated and if system A is more refined than system B, then B's forecasts are equivalent to passing A's forecasts through a noisy filter.

Our fit measure is based on a modified Brier score. Instead of using the predictive probabilities given the data in the computation, we compute the predictive probabilities given the parameters Ω and the data, and use these to compute a Brier score on each

iteration of the Markov chain:

$$BS^{(m)} = \frac{1}{NH} \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{H} \delta_{i,j} \left[z_{i,j,k} - P\left(W_{i,j} = k | \Omega^{(m)}\right)\right]^2$$

where N is the total number of observations; $z_{i,j,k} = 1$ if person i responded k to variable j, and 0 otherwise; and $\delta_{i,j} = 1$ if the variable is observed and 0 if it is missing. That is, missing observations are excluded from the Brier score. The $\{BS^{(m)}\}$ are then used in computing posterior means and standard deviations. This approach extends the Brier score to include calibration, refinement, and uncertainty in the predictive probabilities.

2.6   Model Test on Synthetic Data

Before presenting the results of an empirical application of the model, we discuss the results of a synthetic data analysis mimicking 16 variables measured on a seven point ordinal scale in a sample of 50 persons. The angles θ_j were randomly generated under the constraints of four blocks with four angles in each block. Each item independently had a 7% probability of deletion, and 6.4% of the observations were actually deleted. We estimated circumplex models with the correct constraints on the angles, with unconstrained angles, and with incorrect constraints, assigning variable 4 to block 3 and variable 11 to block 1. The incorrect constraints constitute a mild violation of the true model. We ran the MCMC algorithm for 2000 iterations and used the last 1000 iterations for the analysis. The chains appeared to have converged by iteration 100. Convergence was checked by running longer chains with different starting values and by graphical inspection.

Table 1 reports the posterior mean and RMSSE (root mean squared simulation error) of the log-likelihood and the Brier score. The log-likelihood statistic averages the likelihood function over the posterior distribution of the parameters. It does not play a role in Bayesian inference; however, it is often reported due to its similarity to log-likelihood evaluated at maximum likelihood estimates. The fit statistics identify the correct model. In addition, the unconstrained model yields better fit statistics than the model with the incorrect constraints, even when the incorrect constraints affect only two out of the 16 variables. Note that the

number of parameters are the same for the three models. Using the appropriate constraints improves these fit statistics by reducing the uncertainty in the estimated angles, as will be demonstrated in Table 2. The simulation result lends credence to the use of the Brier score in empirical applications to test the adequacy of constraints on the circumplex.

[INSERT TABLE 1 ABOUT HERE]

The algorithm was able to recover accurately the grand means μ_j, error standard deviations σ_j, and random effects standard deviations λ and τ as judged by their posterior means and posterior standard deviations. We estimated the latent variables $y_{i,j}$, ϕ_i, α_i, and β_i by their posterior means. Their correlations with their true values over "subjects" exceeded 0.95, except in the incorrectly constrained model for $y_{i,11}$ with a correlation of 0.89. The estimated cutpoints had correlation between 0.77 and 0.91 with their true values.

The differences between the models with various sets of constraints are also reflected in the posterior means and standard deviations of the angles in the Table 2. Although the posterior means for the two models are close to their true values, as measured by their posterior standard deviations, the unconstrained model exhibits more uncertainty about the angles, and their posterior variances are larger. Also, the simulation standard errors are substantially smaller for the correctly constrained model. For the unconstrained model the RMSSE is in the range of 0.014 to 0.023; for the correctly constrained model they vary from 0.009 to 0.016, and for the incorrectly constrained model, they are between 0.009 and 0.022, except for item 11, a wrongly classified item, where RMSSE is 0.198. The constraints effectively reduce posterior uncertainty about the angles, as can be seen by comparing the posterior standard deviations of the constrained and unconstrained models. In all cases but one, the posterior distributions were "mound" shaped and unimodal. The exception was item 11 in the incorrectly constrained model where the angle wrapped around the circle. It is worthwhile noting that the incorrectly constrained model did not recover the true angles for the misclassified variables 4 and 11.

[INSERT TABLE 2 ABOUT HERE]

3 Application to Schwartz Value Ratings

3.1 Schwartz Value theory

Building on work by Rokeach (1973), Schwartz and Bilsky (1987, 1990) provided a detailed psychological theory of value content and structure. Values are defined as beliefs that pertain to desirable states or behaviors, transcend specific situations, guide selection or evaluation of behavior and are ordered by relative importance. They have been widely used in the social sciences to explain, for example, voting behavior (Rokeach 1973), mass media usage (Rokeach and Ball-Rokeach 1989), charity contributions (Manner and Miller 1978), socially responsible behavior (Anderson and Cunningham 1972), ecological behavior (Ellen 1994), and innovativeness (Steenkamp, ter Hofstede and Wedel 1999).

Schwartz and his collaborators postulate a comprehensive typology of the content and structure of domains of values. They distinguish ten value domains, summarized in Table 3, that are organized along two bipolar dimensions. The first dimension is defined as "openness to change versus conservation", and opposes values of self-direction and stimulation to those of security, conformity and tradition. The second dimension is called "self-enhancement versus self transcendence" and opposes values of universalism and benevolence to those of hedonism, power, and achievement. The universal structure of values was investigated by Schwartz in a number of studies conducted in different countries. Smallest space analysis (Guttman 1968) of the correlation matrices provided qualitative support for the postulates of the theory (Schwartz and Sagiv 1995). From those studies, it is apparent that the theory borrows strength from ideas of circumplex representations. However, the circumplex model, although popular in psychology as a model for the representation of attitudes (e.g., Plutchik & Conte, 1997) has not been directly applied to the analysis of values, nor has Schwartz theory been subjected to statistical testing. Here we set out to examine more rigorously the validity of Schwartz's value system by investigating the constraints that the theory imposes on the hypothesized circumplex structure.

3.2 Sample and Data

The data used here are part of a larger data set that was collected for the European Commission. A sample was drawn randomly from the household consumer panel of a market research agency in the Netherlands. This panel is representative of the Dutch population with respect to a large number of socio-demographic characteristics. For data collection, mail questionnaires were sent out to households in the Netherlands. The questionnaires included the Schwartz values measurement instrument, assessing 44 value priorities on 9-point scales. Before collecting the data extensive pre–tests were conducted. After sending reminders, the overall response was around 70 percent (for more details on data collection, see Ter Hofstede, Steenkamp and Wedel 1999). The sample size was 157 for a total of 6,698 observations with 3.0% missing.

3.3 Results

We analyzed the Dutch value priority data with seven models derived from Equation (2). The first model has random scale-usage effects only and zeros-out the circumplex structure ($\alpha_i = \beta_i = 0$). We estimated four different circumplex models ($\tau_a = \tau_b = \tau$). Model 2 is an unconstrained circumplex without scale-usage effects ($\phi_i = 0$). Models 3 to 5 are circumplex models with scale-usage effects. Model 3 has value domain constraints, and Model 4 has value dimension constraints. (See Table 3). The latter is a weaker set of constraints as compared to the former. Model 5 is an unconstrained circumplex model. Model 6 is a two–factor model ($\tau_a \neq \tau_b$) that has a more general correlation structure than the circumplex. It also has random scale-usage effects. Model 7 allows an unconstrained error covariance ($\mathrm{cov}(\epsilon_{i,j}, \epsilon_{i,j'}) = \sigma_{j,j'}$) among items and removes the random scale-usage effects and circumplex ($\phi_i = \alpha_i = \beta_i = 0$). The likelihood function with full error covariance and random scale-usage or circumplex factors is not identified. All models were estimated

using 100,000 draws, with a burn-in of 50,000. After burn-in, every 10th iteration was used in the computations for a total of 5000 draws.

Table 4 displays the fit statistics and the root mean squared simulation errors. The latter are quite small, which indicates the accuracy of our model comparisons. The circumplex models with random scale-usage effects (Models 3, 4, and 5) perform better than the other models, and the models without scale-usage effects (Models 2 and 7) performed substantially worse than the models with them. Model 7 is the most general model, but did not perform best: it has a very large number of parameters relative to the sample size. More interestingly, the circumplex models with scale-usage effects (Models 3, 4, and 5) fit as least as well as the two-factor model (Model 6), which is more general than Models 3 to 5 because it relaxes the restriction of equal variance parameters for the factor scores. These fits provide strong evidence that Schartz's value scales conform to the circumplex once scale-usage effects are properly handled. Apparently, the proposed approach of dealing with the response scale bias is both effective and important. For the circumplex models with different constraints the Brier scores are comparatively close. However, the fit deteriorates slightly as domain constraints are imposed. Thus, the value priority data seem to violate Schwartz's theory of value domains to a certain extent, although judged by the differences in fit between those models, the violations appear to be minor.

[INSERT TABLE 4 HERE]

Table 5 reports the posterior means and posterior standard deviations of the estimated circumplex angles for Models 3, 4, and 5, the circumplex models with scale-usage random effects and domain constraints (Full), dimension constraints (Partial), and no constraints (None), respectively. The root mean squared simulation standard errors are small in all cases: they vary from 0.002 to 0.005 for Full and Partial and from 0.008 to 0.010 for None. These RMSSE imply that the numerical approximation accuracy of the posterior mean is within at least two decimal points. Some of the posterior distributions for the angles are bimodal because the support of the distribution spans zero. For example, if there are

16

constraints, angles in the first block can be less than zero. In these cases, we compute the posterior means and standard deviations by "unrolling the circle" in postprocessing the MCMC draws. If the posterior distribution of θ_j is bimodal and if more than half of the posterior distribution is between 0 and π, we recode MCMC draws $\theta_j^{(g)}$ for the gth iteration that are between π and 2π as $\theta_j^{(g)} - 2\pi$. Similarly, if more than half of the distribution is between π and 2π, we recode draws that are between 0 and π as $2\pi + \theta_j^{(g)}$. None of these variables had random deviates in the $\pi/2$ to $3\pi/2$ range. This recoding does not change the circumplex variances and covariance and is only used in approximating the posterior means and standard deviations. Ignoring the bimodal distributions results in nonsensical posterior means and standard deviations: if the posterior distribution is concentrated on both sides of 0, then posterior mean will be around π, a region of zero probability.

When comparing the models with the domain (Full) and the dimension (Partial) constraints, it is apparent that only a few value angles, using the more general dimension constraints, differ from the more exacting ordering, using domain constraints. The violations to the constraints misplace the angles in neighboring value domains within the value dimensions (Table 5). Most of the violations using no constraints, compared to the domain and dimension constraints, occur for angles near zero or 2π.

The posterior standard deviations indicate the uncertainty about the angles. Based on their posterior means and standard deviations, the posterior distribution of the angles from the three models are very similar, with the exceptions of angles for Conservation. Even here, though, the difference are more apparent than real if one keeps in mind that 0.1 radians is very close to 2π on the circle. In comparing posterior standard deviations and simulation standard errors, estimators using domain constraints are more precise than using dimension constraints, which are, in turn, more precise than those without constraints.

[INSERT TABLE 5 HERE]

Figure 2 provides a graphical display of the posterior means for each of the three circumplex models with random intercepts as well as of the model with fixed intercept. The

17

radii of the vectors for angles in the four dimensions have been jittered so that the points do not overlap. Figure 2A shows significant distortions of the circumplex structure if scale usage is not taken into account: the value angles are almost entirely confined to the positive quadrant. Figure 2B graphs the angles for the unconstrained circumplex model with random effects, and Figures 2C and 2D graphs the angles with bipolar dimension and value domain constraints.

Inspection of Figure 2 reveals that, although the unconstrained circumplex does fit the data marginally better, the differences in the locations of the values on the circumplex are minor. To reveal their correspondence, Figure 3 presents scatter plots of the angles for the four circumplex models. Figure 3A plots the angles from the unconstrained circumplex models with and without scale-usage effects and reinforces the observation that scale-usage effects are needed. Figures 3B and 3C plot the constrained models against the unconstrained one with scale-usage effects. The points in the top left-hand corners are due to values in the Conservation dimension overlapping with those of Self-Transcendence in the unconstrained model. Finally, Figure 3D plots the domain and dimension constrained models. These plots indicate that the three circumplex structures are quite similar: the estimated angles are virtually on a straight line. Because of the few and minor violations of the dimension and domain constraints, we are inclined to conclude that the Schwartz value theory holds fairly well in the Netherlands sample, even though the Bayesian model selection criterion points towards the unconstrained model.

[INSERT FIGURES 2 and 3 HERE]

Individual differences in the value judgments are depicted in Figure 4. This figure is based on the bipolar dimension constraints and displays average interaction effects. The averages are over items from the same value dimension:

$$\frac{\alpha}{\text{card}(\mathcal{B}_k)} \sum_{j \in \mathcal{B}_k} \sin(\theta_j) + \frac{\beta}{\text{card}(\mathcal{B}_k)} \sum_{j \in \mathcal{B}_k} \cos(\theta_j) \tag{5}$$

where $\text{card}(\mathcal{B}_k)$ is the cardinality of $\mathcal{B}_k$. Figure 4 contains four curves with α and β equal to plus and minus one. The figure illustrates that a person who has high values for self–

transcendence has low values for self–enhancement and moderate values for openness–to–change and conservation. Likewise, a person with high values for openness–to–change has low values for conservation. Similar patterns can be observed for the other two value dimensions.

[INSERT FIGURES 4 HERE]

4 Conclusion

Models for covariance structures are popular in the social sciences for assessing latent psychological constructs from proxy variables that are intended to represent the psychological domains in question. Whereas the exploratory factor analysis model has been used frequently since 1960, confirmatory factor models (Joreskog 1974) became popular in the 1970's for applications where prior theory guided the identification of the underlying latent variable structure. However, because of their linear form, these broad modelling frameworks can include only a subset of relevant models for covariance structures. One of the significant exceptions, which is not included in the confirmatory factor modelling framework, is the class of circumplex models (Guttman 1954, Browne 1992) which imposes non–linear constraints on the correlation matrix. These constraints are derived from the ordering of the proxy variables on the circumplex and, thus, avoid the need to achieve simple structure through either rotation or identifying constraints.

In our Bayesian formulation of circumplex models for rating scales, we explicitly account for idiosyncratic response scale usage by using an individual level cutpoint approach that assumes that respondents map an underlying latent trait onto the response scale and by a random effects specification that allows for differential scale usage tendencies. In the synthetic data application, we demonstrated that the individual–level cutpoints can be recovered well even when the sample size is small, while the empirical application showed that failure to accommodate response scale usage seriously distorts the recovered circumplex structure. A potential drawback of our Bayesian approach, however, is that as yet standard software is not available and that it requires more computer time than maximum likelihood methods.

19

The circumplex model has been of much appeal to social science researchers because of its implied properties for the correlation structure of the measured items. Our approach yields a tractable representation that deals with different sources of person–specific heterogeneity. Moreover, the Bayesian formulation of the model and the proposed MCMC algorithm allow us to impose inequality constraints on the circumplex that are derived from substantive theory. In the synthetic data analysis and empirical application we showed how to investigate the validity of these constraints. We believe that these contributions will facilitate rigorous tests and further increase the popularity of circumplex models for the analysis of psychological constructs in the social sciences.

A MCMC

All of the full conditionals, except those for the angles, are standard distributions. The MCMC algorithm proceeds by drawing recursively from the full conditional distributions of the parameters, as provided below. Each of those full conditional distributions takes a standard form, with the exception of the full conditional for the angles, θ. We will use the matrix notation and distributions in Section 2.3. The algorithm was implemented in the GAUSS language, and the code can be obtained from the first author.

A.1 Full Conditional for $Y_{i,j}$ for Observed $W_{i,j}$

$$[y_{i,j}|\text{Rest}] \propto N_1[y_{i,j}|\mu_j + \phi_i + \alpha_i \sin(\theta_j) + \beta_i \cos(\theta_j), \sigma_j^2] \, \chi(c_{i,w_{i,j}-1} < y_{i,j} \leq c_{i,w_{i,j}}),$$

where $\chi(\bullet)$ is the indicator function. The full conditional distribution is a truncated normal where the truncation depends on the cutpoints and the observed ordinal response. We use the inverse cdf transform to generate truncated normal random variables.

A.2 Full Conditional for $Y_{i,j}$ for Missing $W_{i,j}$

The model easily accommodates missing data assuming MAR. If $W_{i,j}$ is missing, then $Y_{i,j}$ is normal:

$$[y_{i,j}|\text{Rest}] = N_1[y_{i,j}|\mu_j + \phi_i + \alpha_i \sin(\theta_j) + \beta_i \cos(\theta_j), \sigma_j^2].$$

That is, one does not know which cutpoints $y_{i,j}$ would have fallen between.

A.3 Full Conditional for cutpoints c_i

Given $c_{i,k-1}$, $c_{i,k+1}$, and the latent variables Y_i, the cutpoint $c_{i,k}$ is uniformly distributed:

$$[c_{i,k}|\text{Rest}] = U\left(c_{i,k}\Big| \max_{w_{i,j}=k}(y_{i,j}, c_{i,k-1}), \min_{w_{i,j}=k+1}(y_{i,j}, c_{i,k+1})\right) \text{ for } k = 2, \ldots, H-2.$$

A.4 Full Conditional for $\underline{\mu}$

$$[\underline{\mu}|\text{Rest}] = N_J(\underline{\mu}|\underline{m}_n, V_n)$$

$$V_n = \left(n\Sigma^{-1} + V_0^{-1}\right)^{-1}$$

$$\underline{m}_n = V_n\left(\Sigma^{-1}(Y - \underline{\phi}1_J' - \underline{\alpha}X_S' - \underline{\beta}X_C')'1_n + V_0^{-1}\underline{m}_0\right)$$

A.5 Full Conditional for $\underline{\phi}$

$$[\underline{\phi}|\text{Rest}] = N_n(\underline{\phi}|\underline{m}_\phi, v_\phi I_n)$$

$$v_\phi = \left(1_J'\Sigma^{-1}1_J + \lambda^{-2}\right)^{-1}$$

$$\underline{m}_\phi = v_\phi(Y - 1_n\underline{\mu}' - \underline{\alpha}X_S' - \underline{\beta}X_C')\Sigma^{-1}1_J$$

A.6 Full Conditional for $\underline{\alpha}$ and $\underline{\beta}$

$$[\alpha_i, \beta_i|\text{Rest}] = N_2[(\alpha_i, \beta_i)'|\underline{m}_{\alpha_i,\beta_i}, V_{\alpha_i,\beta_i}]$$

$$V_{\alpha_i,\beta_i} = \left(X'\Sigma^{-1}X + \tau^{-2}I_2\right)^{-1}$$

$$\underline{m}_{\alpha_i,\beta_i} = V_{\alpha_i,\beta_i}X'\Sigma^{-1}(\underline{Y}_{i,\bullet} - \underline{\mu} - \phi_i 1_J)$$

A.7   Full Conditional for σ_j^2

$$[\sigma_j^2|\text{Rest}] = IG\left(\sigma_j^2|\frac{r_n}{2}, \frac{s_n}{2}\right)$$

$$r_n = r_0 + n$$

$$s_n = s_0 + \sum_{i=1}^{n}\left[y_{i,j} - \mu_j - \phi_i - \alpha_i\sin(\theta_j) - \beta_i\cos(\theta_j)\right]^2$$

A.8   Full Conditional for λ^2

$$[\lambda^2|\text{Rest}] = IG\left(\lambda^2|\frac{u_{n,1}}{2}\frac{v_{n,1}}{2}\right)$$

$$u_{n,1} = u_{0,1} + n$$

$$v_{n,1} = v_{0,1} + \underline{\phi}'\underline{\phi}$$

A.9   Full Conditional for τ^2

$$[\tau^2|\text{Rest}] = IG\left(\tau^2|\frac{u_{n,2}}{2}\frac{v_{n,2}}{2}\right)$$

$$u_{n,2} = u_{0,2} + 2n$$

$$v_{n,2} = v_{0,2} + \underline{\alpha}'\underline{\alpha} + \underline{\beta}'\underline{\beta}$$

A.10 Full Conditional for $\underline{\theta}$

$$[\underline{\theta}|\mathrm{Rest}] \propto \exp\left\{-\sum_{i=1}^{n}\sum_{j=1}^{J}\frac{1}{2\sigma_j^2}\left[y_{i,j}-\mu_j-\phi_i-\alpha_i\sin(\theta_j)-\beta_i\cos(\theta_j)\right]^2\right\}\chi(\theta\in\mathcal{C})$$

$$\times \exp\left\{-\frac{1}{2}\sum_{j=2}^{J}[\xi(\theta_j)-\underline{d}_0]'Q_0[\underline{\xi}(\theta_j)-\underline{d}_0]\right\}$$

$$\propto \prod_{j=2}^{J}VM(\theta_j|\underline{d}_j,Q_j,\mathcal{C}_j)$$

$$Q_j = \sigma_j^{-2}\left[\begin{array}{cc}\underline{\alpha}'\underline{\alpha} & \underline{\alpha}'\underline{\beta} \\ \underline{\beta}'\underline{\alpha} & \underline{\beta}'\underline{\beta}\end{array}\right]+Q_0$$

$$\underline{d}_j = Q_j^{-1}\left\{\left[\begin{array}{c}\underline{\alpha}' \\ \underline{\beta}'\end{array}\right](\underline{Y}_{\bullet,j}-\mu_j\underline{1}_n-\underline{\phi})+Q_0\underline{d}_0\right\}$$

where $\mathcal{C}_j$ is the constraint set for θ_j, and $\mathcal{C} = \cup_{j=2}^{J}\mathcal{C}_j$. Because the prior distribution for the angles are independent, the full conditionals depend on each other only through the constraints. Thus, generating from the full conditionals involves generating from univariate distributions, which we do by using the inverse cumulative distribution transform of a uniform random deviate.

Our experience has been that once the random coefficients and angles are in the vicinity of their true values, generating directly from the full conditionals works very well. However, the algorithm can get "stuck", because the random coefficients (X matrix in the posterior density) limit the range of values attainable for the angles and visa versa. Thus, given one set of parameters, draws from the other set may not visit high probability areas of the parameter space.

Therefore, we pursue a hybrid sampling strategy and in addition to generating the angles form their full conditional distributions directly, we generate all the angles and random coefficients in a Metropolis step at every other iteration. The Metropolis step works as follows. We generate the angles from a random walk. The jump distribution is a finite mixture of L uniform distributions where the end points depend on the current angle and the

constraints. Figure 5.a graphs the density of a typical mixture of uniforms, centered at zero, for the error distribution. It puts most mass around zero, which implies that most candidates values, ψ_j, will be close to the current value, θ_j, and it allows large jumps with relatively low probability. Figure 5.b graphs the jump distribution given the current θ_j is 2.5 and the block constraints imply that $0.5 < \psi_j < 2.8$. With probability p_l, the candidate value, ψ_j, for θ_j is generated from a uniform that is proportional to $\chi(\theta_j \in \mathcal{C}_j) U(\psi_j | \theta_j - u_l, \theta_j + u_l)$ where u_l is a prespecified positive constant, and $\mathcal{C}_j$ are the constraints on θ_j given the other values of θ. Once we have candidate values for the angles, we generate candidate values for the random coefficients α and β. The candidate values for the angles and random coefficients are jointly accepted or rejected.

[INSERT FIGURE 5 ABOUT HERE]

The constraints on the angles result in relatively complex expressions, though easy to compute, for the acceptance probabilities. Without loss of generality, suppose $\theta_1 = 0$; $0 < \theta_2 < \pi$; θ_1 and θ_2 are in the first block $\mathcal{B}_1$, and there are K blocks. The indicies $b_1, \ldots, b_K$ will give the last angle that belongs to the blocks:

$$1, \ldots, b_1 \in \mathcal{B}_1 \text{ and } b_{k-1} + 1, \ldots, b_k \in \mathcal{B}_k \text{ for } k = 2, \ldots, K.$$

The blocks follow the order in Section 2.2.

Candidate values $\psi_2, \ldots, \psi_M$ for the angles are generated sequentially. We will use the definitions of the minimum and maximum angles, $(\underline{\mathcal{B}_k}, \overline{\mathcal{B}_k})$, from Section 2.2 where it is to be understood that these minimum and maximum angles change as current values of θ_m are replaced by the candidates ψ_m as the candidates are generated. Define "$\vee$" as the maximum operator, and "$\wedge$" as the minimum operator. The random walk is a mixture of L uniform distributions: in Figure 5a the endpoints for component l are $\pm u_l$ with mixture probability

24

p_l. For angles in the first block, the jump distributions are:

$$g_1[\psi_2|\theta_2,\dots,\theta_J] = \sum_{l=1}^{L} p_l \frac{\chi(0 < \psi_2 < [\underline{\mathcal{B}}_2 \wedge (\theta_2 + u_l) \wedge \pi])}{[\underline{\mathcal{B}}_2 \wedge (\theta_2 + u_l) \wedge \pi]}$$

$$g_1[\psi_m|\psi_2,\dots,\psi_{m-1},\theta_m,\dots,\theta_J] = \sum_{l=1}^{L} p_l \frac{\chi(a_{l,m,1} < \psi_m < b_{l,m,1})}{b_{l,m,1} - a_{l,m,1}}$$

$$a_{l,m,1} = (\overline{\mathcal{B}}_K - 2\pi) \vee (\theta_m - u_l)$$

$$b_{l,m,1} = \underline{\mathcal{B}}_2 \wedge (\theta_m + u_l)$$

$$\text{for } m = 3,\dots,b_1.$$

If the candidate value ψ_m in the first block is negative, then it is recoded as $2\pi + \psi_m$.

For angles in blocks $k = 2$ to K, the uniform random walk has density:

$$g_k[\psi_m|\psi_2,\dots,\psi_{m-1},\theta_m,\dots,\theta_J] = \sum_{l=1}^{L} p_l \frac{\chi(a_{l,m,k} < \psi_m < b_{l,m,k})}{b_{l,m,k} - a_{l,m,k}}$$

$$a_{l,m,k} = \underline{\mathcal{B}}_{k-1} \vee (\theta_m - u_l)$$

$$b_{l,m,k} = \overline{\mathcal{B}}_{k+1} \wedge (\theta_m + u_l)$$

$$\text{for } m = b_{k-1}+1,\dots,b_k.$$

where $\overline{\mathcal{B}}_{K+1} = \underline{\mathcal{B}}_1$.

After generating the candidate angles $\{\psi_m\}$, candidate values of the random coefficients $\{\alpha^c, \beta^c\}$ are generated from normal distributions in Sections A.6. The acceptance probability for the candidates are:

$$\min\left\{1, \frac{[\psi,\alpha^c,\beta^c|Y]\prod_{k=1}^{K}\prod_{m=b_{k-1}+1}^{b_k} g_k[\theta_m|\theta_2,\dots,\theta_{m-1},\psi_m,\dots,\psi_J][\alpha,\beta|\theta]}{[\theta,\alpha,\beta|Y]\prod_{k=1}^{K}\prod_{m=b_{k-1}+1}^{b_k} g_k[\psi_m|\psi_2,\dots,\psi_{m-1},\theta_m,\dots,\theta_J][\alpha^c,\beta^c|\psi]}\right\}$$

where $b_0 + 1 = 2$, and $[\theta,\alpha,\beta|Y] \propto [Y|\mu,\phi,\alpha,\beta,\theta,\sigma][\theta][\alpha][\beta]$.

In the model without block constraints, mixing is improved if one of the uniform distributions in the mixture allows for reflections such as $\chi(\mathcal{C}_j)U(\psi_j|\pi - \theta_j - u_l, \pi + \theta_j + u_l)$ for $j = 2$ and $\chi(\mathcal{C}_j)U(\psi_j|2\pi - \theta_j - u_l, 2\pi + \theta_j + u_l)$ for $j > 2$. The rational for the reflection is that the random walk chain has to progress from a region of high probability, through a region of low probability, to arrive at another area of high probability. For example, suppose

that the current value of the angle is 0.2, so cos(.2) is close to one and sin(.2) is close to zero. Angles close to $2\pi - .2$ result in similar values for the sine, cosine, and covariances among Y variables. However, for a random walk to reach $2\pi - .2$, it has to pass through regions around π where the sines, cosines, and covariances are very different. If .2 is a highly probably value for the angle, the random walk will not reach the other side of the circle because of the low probability region that intervenes. We also included random phase shifts in the algorithm, by adding a small random amount to each angle, which slightly rotates the entire configuration and helps to escape from regions of low probability.

Our experience has been that the random walk Metropolis explores the parameter space more rapidly than generating angles from their full conditional distributions only. However, once the chain is in a high probability region, generating from the full conditionals is more efficient because none of the random deviates are rejected. We therefore perform one random walk Metropolis and one draw from the full conditionals of the angles at each iteration of the Markov chain.

B References

Anderson T.W. (1960). Some stochastic process models for intelligence test scores. In: K.J. Arrow, S. Karlin and P. Suppes (eds.), *Mathematical Methods in the Social Sciences.* Stanford CA: Stanford University Press, 205-220.

Anderson, T.W. and Cunningham, W.H. (1972). Socially conscious consumer behavior, *Journal of Marketing*, 36, 22-31.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability, *Monthly Weather Review,* 78, 1–3.

Browne, M. (1992). Circumplex models for correlation matrices. *Psychometrika*, 57, 469-497.

Cudeck, R. (1986). A note on structural models on the circumplex. *Psychometrika*, 51,

143-147.

Ellen, P.S. (1994). Do we know what we need to know? Objective and subjective knowledge effects on pro-ecological behaviors, *Journal of Business Research*, 30, 43-52.

Gordon, M. D. and Lenk, P. (1991). A utility theoretic examination of the probability ranking principle in information retrieval, *Journal of the American Society for Information Science* 42, 703–714.

Gordon, M. D. and Lenk, P. (1992). When is the probability ranking principle sub–optimal?, *Journal of the American Society for Information Science* 43, 1–14.

Guttman, L. (1954). A new approach to factor analysis: The radex. In: P.F. Lazarsfeld (ed.), *Mathematical Thinking in the Social Sciences*, New York, Columbia University Press, 258-348.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 33, 469-506.

Jöreskog, K.G. (1974). Analyzing psychological data by structural analysis of covariance matrices. IN; D.H. Krantz, R.C. Atkinson, R.D. Luce and P. Suppes (eds.), *Contemporary Developments in Mathematical Psychology*. San Francisco: Freeman, Vol 2, 1-56.

Lippa R. (1995). Gender related individual differences and psychological adjustment in terms of the big five and circumplex models. *Jjournal of Personality and Social Psychology*, 69, 1184-1202.

Manner, L. and Miller, S.J. (1978). An examination of the value-attitude structure in the study of donor behavior. In: *Proceedings of the American Institute of Decision Sciences,* 12, St. Louis, 532-538.

Plutchik, R. and Conte, H. R. (1997). *Circumplex Models of Personality and Emotions.* Washington, DC: American Psychological Association

Rokeach, M.J. (1973). *The Nature of Human Values.* New York, NY: The Free Press.

Rokeach, M.J. and Ball-Rokeach, S.J. (1989). Stability and change in American value priorities, 1968-1981. *American Psychologist*, 44, 775-784.

Rossi, P. E., Gilula, Z. and Allenby, G. M. (2001). Overcoming Scale Usage Heterogene-

ity: A Bayesian Hierarchical Approach. *Journal of the American Statistical Association*, 96, 20-31.

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, 125, 3–30.

Schwartz S.H. and Bilsky, W. (1990). Toward a theory of the universal content and structure of Values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology*, 58, 878-891.

Schwartz, S.H. and Bilsky, W. (1987). Toward a psychological structure of human values. *Journal of Personality and Social Psychology*, 53, 550-562.

Schwartz, S.H. and Sagiv, L. (1995). Identifying culture specifics in the content and structure of values. *Journal of Cross-Cultural Psychology*, 26, 92-116.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function: Part I. *Psychometrika*, 27(3), 125-140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function: Part II. *Psychometrika*, 27(3), 219-246.

Steenkamp, J-B.E.M., Hofstede, F.Ter, and Wedel M. (1999), A cross-national investigation into the individual and cultural antecedents of consumer innovativeness. *Journal of Marketing*, 63, 55-69.

Ter Hofstede, F., Steenkamp, J.E.B.M., Wedel, M. (1999). International market segmentation based on consumer-product relations. *Journal of Marketing Research*, 36, 1-17.

Wiggins, J.S., Steiger, J.H. and Gaelick, L. (1981). Evaluating circumplexity models in personality data. *Multivariate Behavioral Research*, 16, 263-289.

Table 1. Model fit statistics and correlations for synthetic data

RMSSE is the root mean squared simulation error

	Constraints		
	Correct	None	Incorrect
Log Likelihood: Mean	-220.65	-278.47	-321.97
Log Likelihood: RMSSE	1.10	1.30	1.15
Brier Score: Posterior Mean	0.0106	0.0130	0.0146
Brier Score: RMSSE	0.000052	0.000058	0.000052

Table 2. Estimated directions for the synthetic data. The constraints given to the program were either no constraints, the right constraints, or the wrong constraints.

		Posterior Mean			Posterior STD		
Item	True	None	Right	Wrong	None	Right	Wrong
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.673	0.613	0.664	0.618	0.201	0.113	0.092
3	5.700	5.838	5.831	5.856	0.299	0.087	0.096
4	*0.778*	*0.855*	*0.842*	*1.293*	0.191	0.092	0.113
5	0.811	0.772	0.937	0.814	0.222	0.108	0.109
6	1.030	1.127	1.128	1.147	0.198	0.100	0.105
7	1.096	1.116	1.101	1.165	0.197	0.098	0.106
8	1.116	1.309	1.310	1.254	0.183	0.096	0.111
9	2.332	2.531	2.547	2.507	0.195	0.133	0.128
10	2.577	2.676	2.673	2.648	0.180	0.130	0.128
11	*3.285*	*3.494*	*3.497*	*4.144*[†]	0.177	0.137	2.560
12	4.568	4.693	4.622	4.576	0.235	0.116	0.158
13	4.646	4.702	4.723	4.684	0.209	0.104	0.150
14	4.913	5.081	5.070	5.091	0.192	0.102	0.130
15	5.340	5.487	5.418	5.486	0.182	0.103	0.120
16	5.538	5.792	5.732	5.760	0.145	0.096	0.099

[†] The posterior distribution is bimodal with mass near 0 and 2π. The median is 5.78.

30

Table 3. Dimensions, domains, and values according to Schwartz's theory.

Openness–to–Change		Conservation	
Domain	Values	Domain	Values
Self Direction	Creativity Freedom Independent Curious Choosing Own Goals	Security	Family Security National Security Social Order Cleanliness Reciprocation of Favors
Stimulation	Daring Varied Life Exciting Life	Conformity	Politeness Obedient Self-Discipline Honoring Elders
		Tradition	Humble Accepting Fate Religious Respect for Tradition Moderate

Self–Transcendence		Self–Enhancement	
Domain	Values	Domain	Values
Benevolence	Helpful Honest Forgiving Loyal Responsible	Hedonism	Pleasure Enjoying Life
Universalism	Broad Minded Wisdom Social Justice Equality World Peace World Beauty Unity with Nature Environment	Achievement	Successful Capable Ambitious Influential
		Power	Social Power Authority Wealth

Table 4. Fit statistics for the Schwartz Value Data.

	Number[†] of Parameters	Log Likelihood		Brier Score	
		Posterior Mean	STD Error	Posterior Mean	Simulation RMSSE
Model 1	89^a	-10271	0.367	0.0162	0.00000063
Model 2	133^b	-9912	0.447	0.0774	0.00000311
Model 3	134^c	-9752	0.492	0.0157	0.00000079
Model 4	134^c	-9730	0.382	0.0156	0.00000064
Model 5	134^c	-9693	0.433	0.0156	0.00000069
Model 6	135^d	-9822	1.212	0.0157	0.00000156
Model 7	1034^e	-11743	1.494	0.0880	0.00000980

Model 1: Random scale-usage effects and no circumplex ($\alpha_i = \beta_i = 0$)
Model 2: Circumplex ($\tau_a = \tau_b$), no scale-usage effects ($\phi_i = 0$), no constraints
Model 3: Circumplex ($\tau_a = \tau_b$), random scale-usage effects, domain constraints
Model 4: Circumplex ($\tau_a = \tau_b$), random scale-usage effects, dimension constraints
Model 5: Circumplex ($\tau_a = \tau_b$), random scale-usage effects, no constraints
Model 6: Elliptical model ($\tau_a \neq \tau_b$), random scale-usage effects, no constraints
Model 7: Full error covariance

[†] Counting the number of parameters in Bayesian, random effects models is not straightforward. We do not include the latent variables $Y_{i,j}$, the individual-level cutpoints for the ordinal model, the random effects ($\phi_i, \alpha_i, \beta_i$), nor the prior parameters.
[a] 44 means μ_j, 44 error variances σ_j^2, and random effects variance λ^2.
[b] 44 means μ_j, 44 error variances σ_j^2, 44 angles θ_j, and random coefficient variance τ^2.
[c] Same as [b] plus random effects variance λ^2
[d] Same as [c] plus unique τ_a and τ_b instead of common τ
[e] 44 means μ_j and $44(44+1)/2$ error variance and covariance terms.

Table 5. Value Angles for Netherlands Data using Three Sets of Constraints.
"Full" is ordering of angles based on value domains; "Parital" is ordering of angles based on bipolar dimensions. "None" is unrestricted mode.

Dimension						
Domain		Mean			Posterior STD	
Values	Full	Partial	None	Full	Partial	None
Self-Transcendence						
Benevolence						
Helpful	0.000	0.000	0.000	0.000	0.000	0.000
Honest	0.151	0.160	0.581	0.113	0.117	0.292
Forgiving	0.240	0.388	0.864	0.157	0.224	0.353
Loyal	0.326	0.353	0.842	0.144	0.170	0.298
Responsible	0.240	0.281	0.738	0.154	0.178	0.310
Universalism						
Broadminded	1.430	1.289	1.954	0.180	0.169	0.315
Wisdom	0.906	0.825	1.345	0.249	0.232	0.334
Social Justice	0.512	0.249	0.708	0.151	0.176	0.320
Equality	0.977	0.838	1.276	0.253	0.248	0.348
World Peace	0.560	0.399	0.904	0.163	0.190	0.304
World Beauty	0.739	0.613	1.086	0.203	0.212	0.325
Unity with Nature	0.623	0.443	0.946	0.179	0.203	0.324
Environment	0.952	0.834	1.306	0.208	0.199	0.321
Openness-to-Change						
Self-Direction						
Creativity	2.023	1.964	2.861	0.199	0.208	0.386
Freedom	1.556	1.406	1.577	0.194	0.179	0.324
Independent	1.802	1.576	1.703	0.266	0.256	0.364
Curious	1.945	1.862	2.413	0.209	0.212	0.360
Choosing Own Goals	1.869	1.735	2.129	0.204	0.205	0.317
Stimulation						
Daring	2.274	2.086	3.488	0.173	0.176	0.333
Varied Life	2.225	1.988	2.575	0.173	0.178	0.317
Exciting Life	2.214	1.916	2.430	0.173	0.188	0.317

Table 5 Continued

Dimension						
Domain		Mean			Posterior STD	
Values	Full	Partial	None	Full	Partial	None
Self-Enhancement						
Hedonism						
Pleasure	2.346	2.190	2.076	0.174	0.180	0.331
Enjoyment	2.352	2.202	2.290	0.174	0.179	0.331
Achievement						
Successful	3.088	2.999	3.648	0.221	0.212	0.331
Capable	2.455	2.192	1.663	0.198	0.192	0.329
Ambitious	3.457	3.659	4.327	0.225	0.269	0.365
Influential	3.485	3.608	4.252	0.206	0.233	0.350
Power						
Social Power	3.901	3.779	4.341	0.256	0.253	0.341
Authority	3.920	3.832	4.397	0.230	0.228	0.334
Wealth	3.647	3.166	3.722	0.199	0.242	0.338
Conservation						
Security						
Family Security	5.572	5.833	0.176	0.179	0.220	0.350
National Security	5.536	5.850	0.240	0.201	0.238	0.374
Social Order	5.645	6.084	0.495	0.154	0.142	0.315
Cleanliness	4.989	4.927	5.479	0.220	0.224	0.322
Reciprocation of Favors	4.888	4.922	5.719	0.314	0.350	0.461
Conformity						
Politeness	5.816	5.925	0.192	0.134	0.192	0.334
Obedient	5.798	5.672	6.192	0.133	0.205	0.309
Self-Discipline	5.852	6.153	0.821	0.133	0.113	0.328
Honoring Elders	5.806	5.664	6.151	0.133	0.207	0.315
Tradition						
Humble	6.041	5.348	5.637	0.135	0.341	0.371
Accepting Fate	6.154	6.086	0.743	0.103	0.178	0.367
Devout	5.982	4.743	5.262	0.135	0.279	0.362
Respect for Tradition	5.999	5.489	6.076	0.137	0.273	0.380
Moderate	6.122	6.038	0.475	0.114	0.169	0.315

34

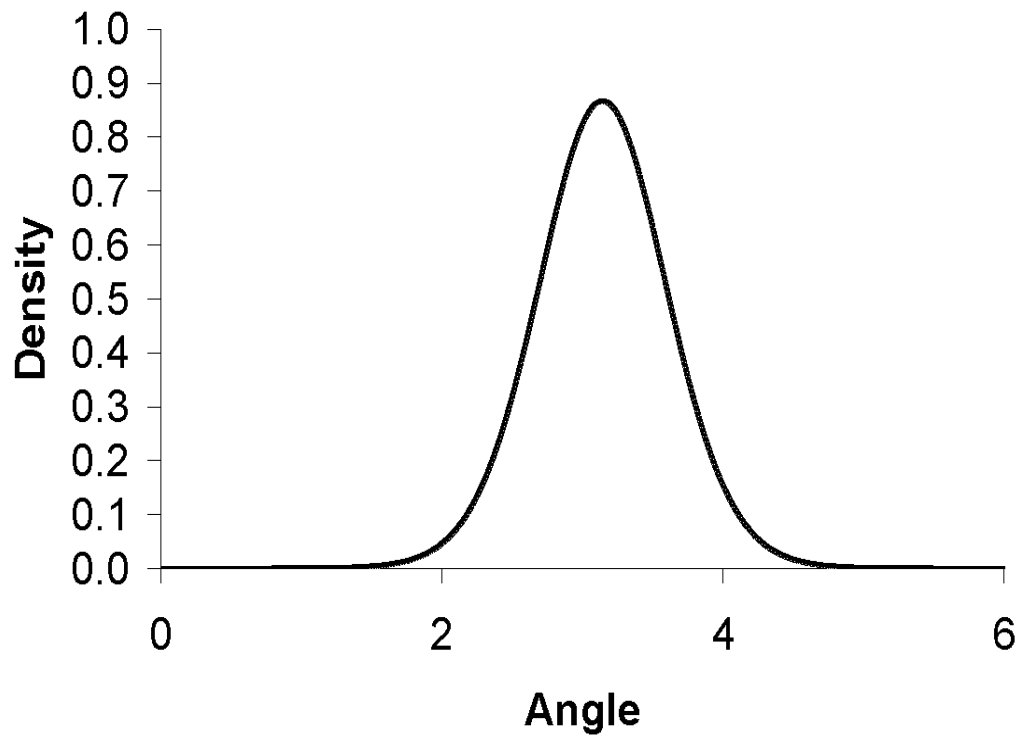Figure 1. Extended Von Mises Distribution.

Figure 2. Sine versus Cosine of posterior means of item angles for Schwartz value data: A. no constraints and fixed intercepts, B. no constraints and random intercepts, C. bipolar constraints and random intercepts, and D. value domain constraints and random intercepts. Angles are identified by their bipolar dimensions: T = self-transcendence, O = openness-to-change, E = self-enhancement, and C = conservation.
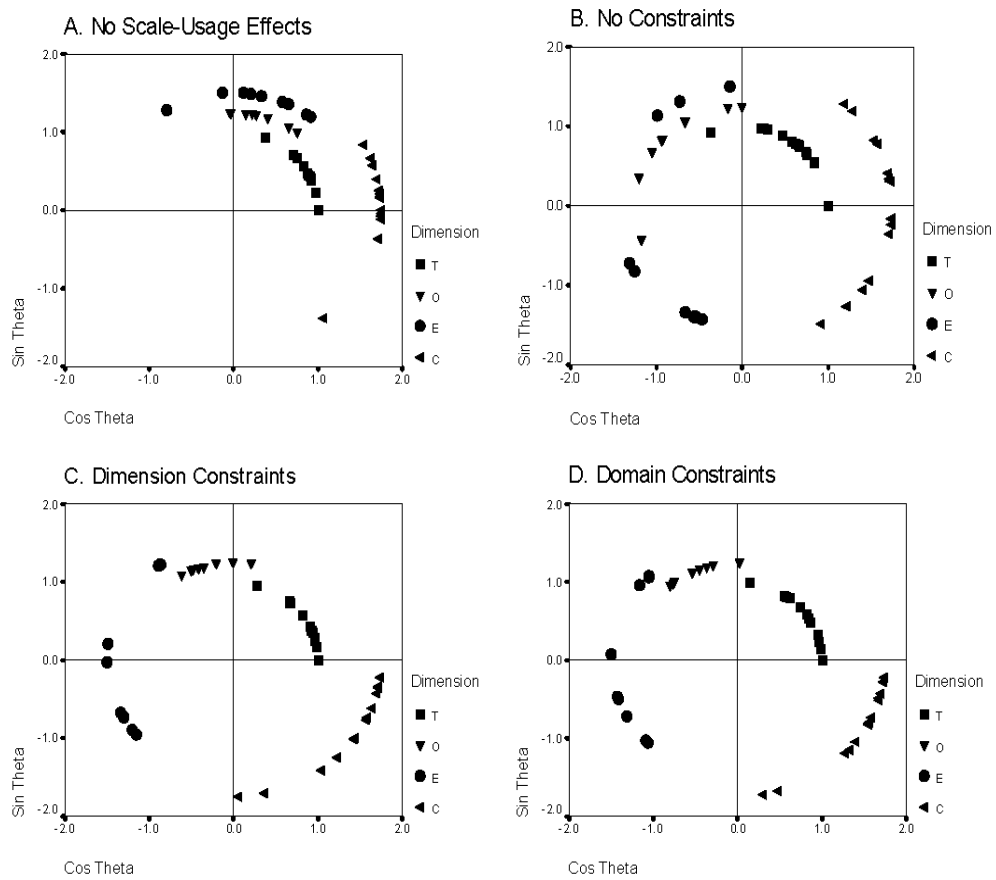
Figure 3. Plots of Angles from different Circumplex Models. A. Scale-usage effects versus no scale-usage effects without constraints. B. Dimension constraints versus no constraints with scale-usage effects. C. Domain constraints versus no constraints with scale-usage effects. D. Domain constraints versus dimension constraints with scale-usage effects. T = self-transcendence, O = openness-to-change, E = self-enhancement, and C = conservation.
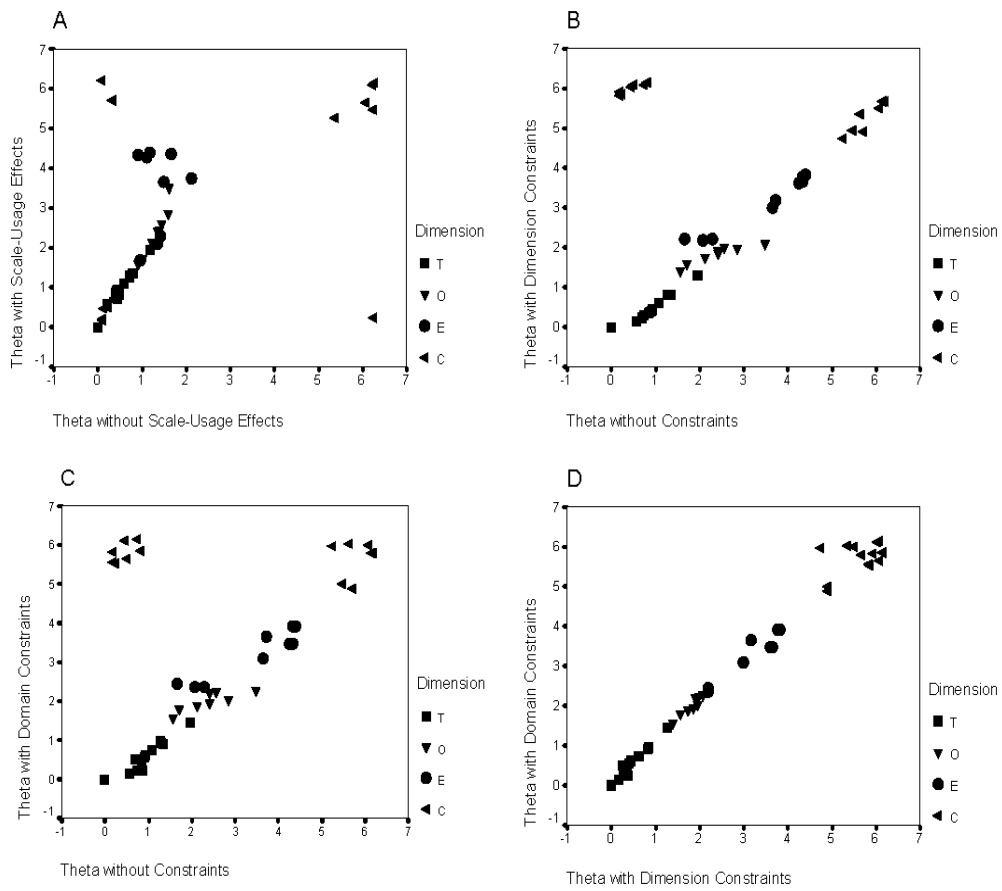
Figure 4. Plot of average interactions for selected values of the random coefficients for the Schwartz value model with bipolar dimension constraints. The four curves are averaged interaction effects where the averaging is within value dimensions for different combinations of $\alpha = \pm 1$ and $\beta = \pm 1$ in Equation (5).
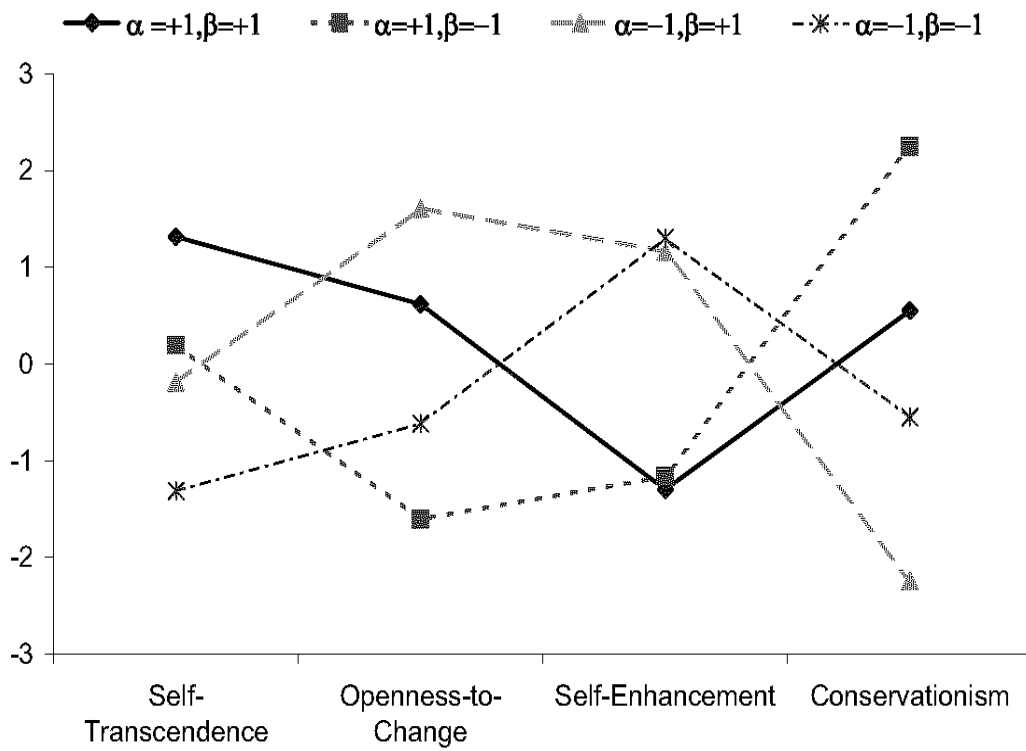
Figure 5. Random walk jump distribution for angles. a. Mixture of four uniforms. b. Random walk based on mixture of uniforms assuming that the current value of the angle is 2.5 radians and the block constraints imply that the angle is restricted between 0.5 and 2.8 radians.